Insights for disease modeling from single cell transcriptomics of iPSC-derived neurons and astrocytes across differentiation time and co-culture.

Das D[1&], Sonthalia S[2&], Stein-O'Brien G[1], Wahbeh M[1], Feuer K[1], Colantuoni C[3,4,5*], Machairaki V[1], and Avramopoulos D[1,6*]

1. Department of Genetic Medicine, Johns Hopkins University School of Medicine
2. Department of Biomedical Engineering, Johns Hopkins University School of Medicine
3. Department of Neuroscience, Johns Hopkins University School of Medicine
4. Department of Neurology, Johns Hopkins University School of Medicine
5. Institute of Genome sciences, University of Maryland School of Medicine
6. Department of Psychiatry, Johns Hopkins University School of Medicine

[&] These authors contributed equally

* Corresponding authors

Mailing address: 733 N. Broadway, MRB-507, Baltimore, MD 21205

Tel. 410 955-8323

Fax: 410 955-7397

Email: adimitr1@jhmi.edu

Keywords: Schizophrenia, autism, Alzheimer's, transcriptome, single cell, induced pluripotent stem cells, neurons, astrocytes, Ngn2.

Running title: iPSC-derived neurons for disease modeling

ABSTRACT

Trans-differentiation of human induced pluripotent stem cells into neurons (hiPSC-N) via Ngn2-induction has become an efficient system to quickly generate neurons for disease modeling and in vitro assay development, a significant step up from previously used neoplastic and other cell lines. Recent single-cell interrogation of Ngn2-induced neurons however, has revealed some similarities to unexpected neuronal lineages. Similarly, a straightforward method to generate hiPSC derived astrocytes (hiPSC-A) for the study of neuropsychiatric disorders has also been described. Here we examine the homogeneity and similarity of hiPSC-N and hiPSC-A to their in vivo counterparts, the impact of different lengths of time post Ngn2 induction on hiPSC-N (15 or 21 days) and of hiPSC-N / hiPSC-A co-culture. We explore how often genes differentially expressed between conditions relate to genetic risk for neuropsychiatric disease. Leveraging the wealth of existing public single-cell RNA-seq (scRNA-seq) data in Ngn2-induced neurons and in vivo data from the developing brain, we provide perspectives on the lineage origins and maturation of hiPSC-N and hiPSC-A. Both show heterogeneity and share similarity with multiple in vivo cell fates, and both cell types more precisely approximate their in vivo counterparts when co-cultured. hiPSC-A show more heterogeneity and similarities to other non-neural cell types, especially when cultured in isolation. Gene expression data from the hiPSC-N show excess of genes linked to schizophrenia (SZ) and autism spectrum disorders (ASD) as has been previously shown for neural stem cells and neurons. These overrepresentations of disease genes are strongest in our system at early times (day 15) in Ngn2-induction/maturation of neurons, which together with our observation of similarities with in vivo neurons earlier in development suggest they may be a better model for neurodevelopmental disorders. We have assembled this new scRNA-seq data along with the public data explored here as an integrated biologist-friendly web-resource for researchers seeking to understand this system more deeply: https://nemoanalytics.org/p?l=DasEtAl2022 .

INTRODUCTION

Recent advances in cell engineering have provided unprecedented tools for investigating the biology and genetics underlying psychiatric disorders [1]. For many years our only opportunity to study the central nervous system (CNS) and create disease models was through model organisms like worms and mice or through CNS and other tumor derived cell lines. These models, while valuable in understanding how the CNS functions, came with significant limitations when drawing parallels to the complex human brain and the related disorders. Four recent technologies have drastically widened the array of tools to model disease: The generation of human induced pluripotent stem cells (hiPSCs) from somatic cells, techniques for differentiation to specific cell types, genome editing, and high throughput transcriptomics including single-cell sequencing (scRNA-seq). We can now generate pluripotent cells from patients or controls, introduce precise genetic modifications that we wish to study into these cells, and generate different types of cells [2-7], such as glutamatergic, GABAergic, Dopaminergic neurons or glial cells. We can study phenotypic consequences using the transcriptome as a readout either in bulk or at single-cell resolution which allows us to detect and account for cellular heterogeneity. With all these advances cellular models are quickly becoming a front-line tool in schizophrenia research. However, there are important limitations to consider when working with such models. Specifically, in the case of two-dimensional (2D) cultures (as opposed to three-dimensional organoids) different neural cell types are most often grown in isolation, in the absence of the milieu of neural types and supporting cells that are found in vivo and play important roles in each other's development and function. While organoid systems allow access to more complex cellular interactions and more advanced maturational states, 2D systems often produce more uniform cell states that are more amenable to assay development for assessing novel therapeutics. Furthermore, the differentiation technologies are far from a recapitulation of in vivo differentiation and although the similarities to the target cell types have been shown [1] it is known that significant differences also exist. scRNA-seq provides increased resolution to answer some key questions on cell type identity and state. Previously we could only study individual cell types through laborious and imperfect approaches such as laser capture microdissection [8]. Today, by acquiring transcriptomes from single cells, either from cultures or from living tissues, we can get a better resolved picture of the component cell types/states of a culture population and perform direct comparisons between in vivo and in vitro differentiated cells.

In this study we focus on human excitatory neurons and astrocytes derived from iPSCs (hiPSC-N and hiPSC-A). We use a transcription factor (Ngn2)-mediated rapid induced differentiation protocol [7] to generate neurons a method that is popular due to its speed and versatility of starting cell type (lymphocytes, fibroblasts, iPSCs etc.) as described [7]. We also use a previously described directed differentiation protocol [9] to generate hiPSC-A. Studying these differentiated cells in 2D cultures can be a powerful approach to model human psychiatric disease [1]. Yet, to best interpret any observed cellular phenotyping results, it is important to test our cells on four different parameters/attributes: (1) How much do these cells resemble the intended cell types in vivo? (2) How homogeneous are they in 2D cultures? (3) How does variation in the differentiation time and co-culture with human astrocytes affect the neural identity of the cells and (4) how similar are cells produced by different laboratories using the

same or similar methods? To help answer these questions we examined by single cell RNA sequencing (scRNA-seq) four conditions: (A) hiPSC-N after 21 days of differentiation (hiPSC-N21) (B) hiPSC-N after 15 days of differentiation (hiPSC-N15) (C) hiPSC-A grown alone (hiPSC-A0) and (D) hiPSC-A co-cultured with hiPSC-N21 (hiPSC-N21A and hiPSC-AN21). We then explored whether transcriptome-based analysis after ScRNA-Seq can distinguish between cells from different conditions, we identified genes with significant differential expression between conditions, investigated their enrichment in functional pathways and in disease associate genes, and explored how these induced neurons and astrocytes compared to other comparable in vivo and in vitro studies.

In this study we refer to multiple external datasets which we examine alongside ours. To provide a single location where all the diverse datasets examined in this report can be accessed and explored in an integrated environment, we have created a web resource leveraging the gEAR and NeMO Analytics platforms[10]. We invite researchers to explore this resource that can visualize individual genes of interest or sets of genes simultaneously across the many datasets used here: https://nemoanalytics.org/p?l=DasEtAl2022 .

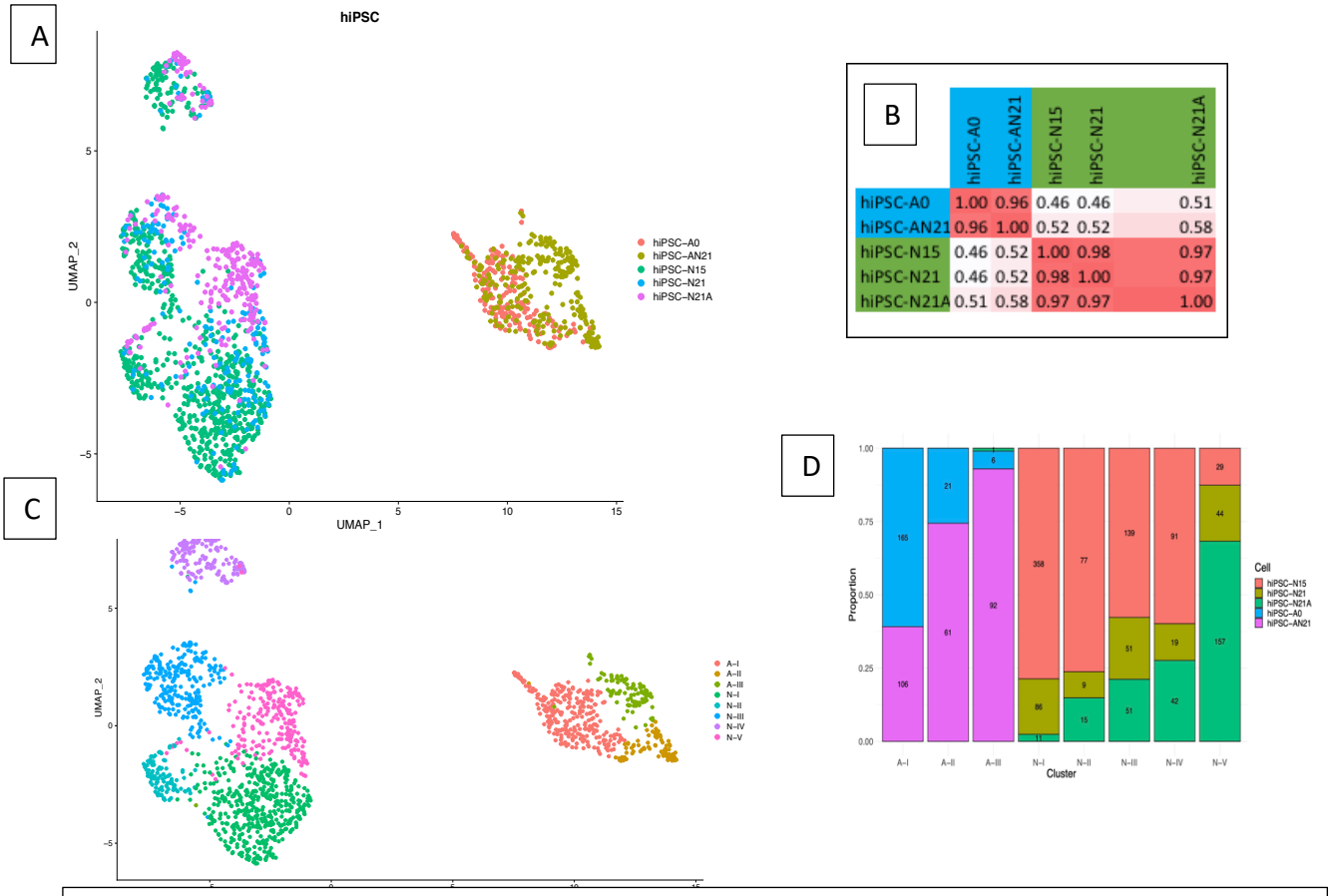

Figure 1: UMAP plot of our data highlighting the different conditions (A), pairwise correlations of gene expression in bulk and pseudobulk gene expression for the conditions (B), he same UMAP plot highlighting cluster derived with Louvain clustering (C) and compositions of the Louvain clusters in terms of cells from the five conditions

RESULTS

1. Cell type marker genes and cellular heterogeneity

We first performed Principal Component Analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction to explore the clustering of cells and homogeneity within cell types. As expected, we observed two clearly separating major clusters corresponding to hiPSC-N and hiPSC-A (Figure 1A). This allowed us to cleanly separate hiPSC-N21A from hiPSC-AN21 which were grown together. We then calculated pair-wise correlations (r^2) of pseudobulk expression of all genes between the 5 conditions: hiPSC-N15, hiPSC-N21, hiPSC-N21A, hiPSC-A0 and hiPSC-AN21 (Figure 1B). Within intended cell types (i.e., within hiPCS-N or hiPCS-A) the correlations between the different conditions were strong (r^2>0.97) while across different intended cell types they were significantly weaker (r^2 < 0.58) further highlighting the difference between hiPSC-N and hiPSC-A. To determine how faithfully each iPSC derived group represents in vivo excitatory neurons and astrocytes we explored the expression of astrocytic and neuronal marker genes. Table 1 shows the expression average across conditions for each gene in CPM (grey shaded column) followed by the standard deviations from this mean for each condition (blue-red shaded column). As expected hiPSC-A0 and hiPSC-AN21 highly expressed astrocytic markers, much more so than neurons, and this was more pronounced in the hiPSC-AN21 which may be an indication of higher maturity. Similarly, in hiPSC-N15, hiPSC-N21 and hiPSC-N21A we observed high expression of many neuronal markers as expected, with a trend for higher expression in the direction hiPSC-N15→hiPSC-N21→hiPSC-N21A especially for glutamatergic marker genes. hiPSC-N also showed expression of other neuronal fate markers.

Few genes did not follow the expected trends, but in most cases, this could be explained by previous published observations. The most striking was the glutaminase gene (*GLS*) encoding an enzyme that generates glutamate from glutamine and often considered a glutamatergic neuronal marker leading us to expect higher expression in hiPSC-N, as opposed to the observed higher expression in hiPSC-A. In agreement with our data however, Cordona et al [11] have shown very high expression of *GLS* in in vivo astrocytes both by immunofluorescence and by RT-PCR. *TPH2, TH* and GAD1 were low, but somewhat higher in expression in hiPSC-A. This higher expression of *GAD1* (GAD67) may not be surprising as it is consistent with the results of Lee et al [12] who showed strong GAD67 staining in astrocytes. We cannot explain the results for TPH2 and TH, though given their low expression levels it may be due to noise. Within astrocytes expression of the 3 genes was lower in hiPSC-AN21 than hiPSC-A0. This, together with the higher expression of astrocytic markers in hiPSC-AN21 may be an additional indication that coculture with hiPSC-N enhances their maturation. While most GABAergic markers were low in hiPSC-N the GABA receptors *GABBR1* and *GABBR1* were high across conditions which is expected for glutamatergic neurons [13] and suggests that the neurons derived by Ngn2 induction may be responsive to GABA. Oligodendrocyte markers *OLIG1* and *OLIG2* were low across conditions but *CNP*, encoding 2',3'-Cyclic Nucleotide 3' Phosphodiesterase, which is an oligodendrocyte marker in the central nervous system and not active in astrocytes [14] was

Table 1

Other names and information	Gene	AVG CPM	Ast	AstN21	N15	N21	N21Ast	cell types
Glial fibrillary acidic protein	*GFAP*	1.05	-0.140	1.694	-0.528	-0.124	-0.903	ASTROCYTES
GLAST/EAAT1	*SLC1A3*	1.18	-0.172	1.737	-0.682	-0.663	-0.220	ASTROCYTES
Vimenttin	*VIM*	885.83	0.404	1.586	-0.683	-0.683	-0.625	ASTROCYTES
Glia-Derived Nexin	*SERPINE2*	68.57	1.564	0.449	-0.683	-0.684	-0.646	ASTROCYTES
paraoxonase 2	*PON2*	50.92	0.433	1.572	-0.679	-0.669	-0.657	ASTROCYTES
Apolipoprotein E	*APOE*	3.44	-0.421	1.786	-0.514	-0.490	-0.362	ASTROCYTES
Neuraminidase 1	*NEU1*	3.47	0.607	1.463	-0.808	-0.779	-0.483	ASTROCYTES
Connexin 43	*GJA1*	14.75	1.201	0.983	-0.734	-0.736	-0.714	ASTROCYTES
late marker of astrocyte development - 17078026	*S100B*	1.46	-0.232	1.685	-0.802	-0.681	0.030	ASTROCYTES
Spondin 2	*SPON2*	1.18	0.062	1.706	-0.695	-0.652	-0.421	ASTROCYTES
ATPase Na+/K+ Transporting Subunit Alpha 2	*ATP1A2*	0.19	0.240	1.515	-1.010	0.046	-0.791	ASTROCYTES
Protein arginine N-methyltransferase 3	*PRMT3*	1.37	0.938	1.175	-0.795	-0.312	-1.005	ASTROCYTES
hevin - produced by astrocytes controls excitatory synaptogenesis	*SPARCL1*	1.11	-0.349	1.782	-0.528	-0.541	-0.363	ASTROCYTES
Glutamate/Aspartate Transporter II	*SLC1A2*	4.57	-1.073	-1.061	0.347	0.841	0.946	GLUTAMATERGIC
Vesicular glutamate transporter 1	*SLC17A7*	1.19	-1.094	-1.094	0.745	0.780	0.664	GLUTAMATERGIC
vesicular glutamate transporter 2	*SLC17A6*	22.37	-1.062	-1.045	1.052	0.286	0.770	GLUTAMATERGIC
Glutaminase	*GLS*	32.88	1.548	0.479	-0.711	-0.687	-0.629	GLUTAMATERGIC
Glutaminase 2	*GLS2*	1.22	-1.134	-0.875	0.271	1.296	0.442	GLUTAMATERGIC
Glutamine synthetase	*GLUL*	4.68	-1.361	-0.547	1.258	0.491	0.159	GLUTAMATERGIC
glutamate ionotropic receptor AMPA type	*GRIA1*	5.66	-0.831	-0.685	-0.429	0.355	1.590	GLUTAMATERGIC
glutamate ionotropic receptor AMPA type	*GRIA2*	11.25	-1.084	-1.054	0.361	0.935	0.842	GLUTAMATERGIC
glutamate ionotropic receptor AMPA type	*GRIA3*	3.23	-0.908	-1.167	0.855	1.021	0.198	GLUTAMATERGIC
glutamate ionotropic receptor AMPA type	*GRIA4*	26.95	-1.073	-1.060	0.701	0.392	1.040	GLUTAMATERGIC
glutamate ionotropic receptor NMDA type	*GRIN1*	1.14	-1.068	-1.068	0.650	1.052	0.434	GLUTAMATERGIC
glutamate ionotropic receptor NMDA type	*GRIN2A*	2.35	-1.330	-0.722	0.368	0.560	1.124	GLUTAMATERGIC
glutamate ionotropic receptor NMDA type	*GRIN2B*	1.40	-0.779	-0.691	-0.289	0.079	1.681	GLUTAMATERGIC
glutamate ionotropic receptor NMDA type	*GRIN2D*	2.27	-1.056	-0.932	0.003	0.849	1.137	GLUTAMATERGIC
glutamate ionotropic receptor NMDA type	*GRIN3A*	2.40	-0.786	-0.711	-0.340	0.187	1.651	GLUTAMATERGIC
HuD - expressed in glutamatergic excitatory neurons	*ELAVL4*	95.17	-1.079	-1.072	0.766	0.427	0.958	GLUTAMATERGIC
Acetylcholinesterase	*ACHE*	6.01	-0.978	-0.837	-0.268	0.815	1.268	CHOLINERGIC
choline acetyl transferase	*CHAT*	1.11	-1.130	-0.982	0.309	1.039	0.764	CHOLINERGIC
vesicular acetylcholine transporter	*SLC18A3*	1.77	-1.185	-0.977	0.761	0.498	0.903	CHOLINERGIC
tryptophan hydroxylase 1	*TPH1*	0.06	-0.994	-0.994	0.735	0.035	1.217	SEROTONINERGIC
tryptophan hydroxylase 2	*TPH2*	0.10	1.022	-1.627	-0.153	0.382	0.376	SEROTONINERGIC
serotonin transporter	*SLC6A4*	0.10	-1.385	-0.309	1.057	-0.259	0.895	SEROTONINERGIC
tyrosine hydroxylase	*TH*	1.05	1.479	-1.070	-0.771	0.245	0.117	DOPAMINERGIC
dopamine transporter (DAT1)	*SLC6A3*	1.00	NA	NA	NA	NA	NA	DOPAMINERGIC
paired like homeodomain 3	*PITX3*	1.00	NA	NA	NA	NA	NA	DOPAMINERGIC
Synaptic vesicular amine transporter - VMAT2	*SLC18A2*	1.18	-0.573	-0.550	-1.027	1.056	1.095	DOPAMINERGIC
GABA transporter type 1	*SLC6A1*	1.01	-1.014	-1.014	1.190	0.717	0.120	GABAERGIC
Glutamate decarboxylase 2 = GAD65	*GAD2*	1.11	-1.221	-0.859	0.470	1.160	0.451	GABAERGIC
Glutamate decarboxylase 1 = GAD67	*GAD1*	1.18	1.333	0.736	-0.915	-0.277	-0.877	GABAERGIC
GABA type B receptor subunit 1	*GABBR1*	3.18	-1.175	-0.822	0.027	0.994	0.976	GABAERGIC
GABA type B receptor subunit 2	*GABBR2*	2.78	-1.115	-0.959	0.289	0.611	1.175	GABAERGIC
2',3'-Cyclic Nucleotide 3' Phosphodiesterase	*CNP*	8.79	1.029	1.130	-0.942	-0.743	-0.474	OLIGODENRD.
Oligodendrocyte Transcription Factor 1	*OLIG1*	1.15	-1.113	-1.003	0.954	0.873	0.290	OLIGODENRD.
Oligodendrocyte Transcription Factor 2	*OLIG2*	0.00	NA	NA	NA	NA	NA	OLIGODENRD.
nestin - neural development	*NES*	3.01	-0.039	1.736	-0.447	-0.657	-0.594	NEURODEV.
neuroectoderm cell fate determinant	*PAX6*	1.33	0.877	1.277	-0.892	-0.574	-0.689	NEURODEV.
neuronal differentiation in cerebellar development,	*FOXG1*	1.02	-0.639	1.735	-0.416	-0.639	-0.042	NEURODEV.
NG2 - OPC marker	*CSPG4*	1.43	1.656	0.240	-0.657	-0.657	-0.582	NEURODEV.
TCIP2 - B cell leukemia 11b	*BCL11B*	1.16	-0.687	-1.267	0.384	0.263	1.307	NEURODEV.
TF neuronal differentiation = BRN2	*POU3F2*	1.59	-0.864	0.075	1.289	0.605	-1.105	NEURODEV.
Trans. Factor, maturation of excitatory cortical neurons.	*FOXP2*	2.26	1.353	0.575	-0.243	-0.420	-1.266	NEURODEV.
Microtubule Associated Protein 2	*MAP2*	331.71	-1.115	-1.056	0.912	0.541	0.718	OTHER NEUR
L1 cell adhesion molecule	*L1CAM*	7.66	-1.063	-1.042	0.292	0.742	1.071	OTHER NEUR
Microtubule Associated Protein Tau	*MAPT*	133.04	-1.081	-1.075	0.594	0.555	1.007	OTHER NEUR
NEUN - Neuronal Splicing Regulator	*RBFOX3*	20.85	-0.983	-0.972	0.128	0.466	1.362	OTHER NEUR
Tubulin Beta 3 Class III (TUJ1)	*TUBB3*	4.42	-0.883	-0.823	0.140	1.588	-0.023	OTHER NEUR
Neurofilament Medium	*NEFM*	283.61	-1.021	-0.987	1.322	0.376	0.310	OTHER NEUR
Neurofilament heavy	*NEFH*	1.16	-0.847	-0.693	-0.077	1.671	-0.055	OTHER NEUR
Neurofilament light	*NEFL*	900.58	-1.102	-1.033	0.596	0.471	1.068	OTHER NEUR
Neuroligin 1 - neuronal cell surface protein	*NLGN1*	81.62	-1.088	-1.089	0.570	0.716	0.891	OTHER NEUR
Transcription factor, regulate dendritic branching, spine	*CUX1*	188.64	-1.161	-0.987	1.021	0.567	0.560	OTHER NEUR
SATB Family Member 2	*SATB2*	5.07	1.277	0.892	-0.753	-0.755	-0.662	OTHER NEUR
maintenance of Purkinje neuron axons	*ELAVL3*	100.33	-1.093	-1.088	0.612	0.694	0.875	OTHER NEUR
neural cell adhesion molecule 1	*NCAM1*	337.66	-1.095	-1.091	0.836	0.651	0.699	OTHER NEUR
doublecortin - neuronal migration, microtubule stabilization	*DCX*	461.32	-1.084	-1.075	0.880	0.442	0.837	OTHER NEUR
enolase 2, Neuron-Specific Enolase	*ENO2*	42.25	-1.073	-1.014	0.494	0.407	1.187	OTHER NEUR
calcium/calmodulin dependent protein kinase 1	*CAMK1*	2.34	-1.654	0.022	0.053	0.874	0.705	CaM Kinases
calcium/calmodulin dependent protein kinase 2A	*CAMK2A*	1.46	-1.045	-0.940	0.183	1.329	0.473	CaM Kinases
calcium/calmodulin dependent protein kinase 2B	*CAMK2B*	19.21	-1.033	-0.985	0.072	0.790	1.156	CaM Kinases
calcium/calmodulin dependent protein kinase 2G	*CAMK2G*	6.31	-0.928	-0.643	-0.572	0.843	1.299	CaM Kinases
calcium/calmodulin dependent protein kinase 2N1	*CAMK2N1*	247.46	-0.981	-0.772	0.100	0.097	1.556	CaM Kinases
calcium/calmodulin dependent protein kinase 2N2	*CAMK2N2*	3.71	-0.909	-0.870	0.123	0.110	1.546	CaM Kinases
calcium/calmodulin dependent protein kinase 4	*CAMK4*	3.30	-0.985	-0.914	0.150	1.442	0.307	CaM Kinases
calcium/calmodulin dependent protein kinase 2D	*CAMK2D*	59.10	0.811	1.337	-0.817	-0.678	-0.653	CaM Kinases
Synaptophysin	*SYP*	26.51	-1.052	-1.022	0.166	0.942	0.967	SYNAPTIC
synapsin 1	*SYN1*	4.32	-1.048	-0.977	0.083	0.807	1.135	SYNAPTIC
synapsin 2	*SYN2*	2.87	-0.933	-0.965	-0.013	0.525	1.386	SYNAPTIC
Synaptotagmin 1	*SYT1*	382.58	-1.062	-1.089	0.922	0.813	0.417	SYNAPTIC
Neurexin 3	*NRXN3*	23.04	-0.841	-0.932	0.353	-0.093	1.514	SYNAPTIC
PSD95	*DLG4*	8.26	-1.006	-0.939	-0.075	1.188	0.831	SYNAPTIC
PSD-localized scaffolding protein, interacts with Shank3. regulates mGluR	*HOMER1*	23.83	-0.972	-1.046	0.443	0.275	1.300	SYNAPTIC
Neurexin 1, excitatory Synaptic marker	*NRXN1*	699.46	-1.092	-1.089	0.869	0.718	0.593	SYNAPTIC
Neurexin 2, Synaptic marker	*NRXN2*	16.98	-1.008	-0.981	-0.019	1.052	0.956	SYNAPTIC
gephyrin - inhibitory postsynaptic scaffold protein,	*GPHN*	203.09	-1.119	-1.065	0.665	0.861	0.657	SYNAPTIC

detected in hiPSC-A indicative of a small deviation from astrocytic identity towards another glial cell type. An array neural development genes were low in all our cells with no clear pattern of differential expression, supporting their maturity. One exception was nestin (NES) a neural progenitor cell marker which is turned off in the mature nervous system [15] and which we found expressed at moderate levels in hiPSC-A0 and hiPSC-AN21. Cho et al [16] have shown high expression of *NES* in mature astrocytes, induced by ischemia in the CA1 hippocampal region. It is therefore possible that the culture conditions also induce *NES* in our astrocytes. An array of other neuronal marker genes showed higher expression in hiPSC-N than hiPSC-A with one exception, *SATB2*. *SATB2* is a postmitotic determinant for upper-layer neuron specification not present in all neurons [17]. While this can explain its absence in hiPSC-N it is not clear why we observe it expressed in hiPSC-A. We examined the expression of multiple Calcium/calmodulin-dependent protein kinases (CaMK) and most were highly expressed in neurons. The only exception was *CAMK2D* whose expression was higher in astrocytes. This is in agreement with Vallano et al [18] who have shown that CAMK2D is a CaM Kinase type II with specific astrocyte expression. Finally, we examined several synaptic markers, all of which showed high expression in hiPSC-N, with a trend for higher expression in the direction hiPSC-N15→hiPSC-N21→hiPSC-N21A like the glutamatergic markers. This positive correlation of synaptic gene expression with time post Ngn2 induction and co-culture with astrocytes indicates increasing maturation of the neurons along the hiPSC-N15→hiPSC-N21→hiPSC-N21A axis.

We next harnessed the power of Single cell sequencing to explore the cellular homogeneity of these hiPSC derived differentiated cells. We applied Louvain community detection, a method to extract communities with shared features from large networks [19], which identified 8 clusters within our cells (Figure 1C). Five clusters included exclusively derived neurons and we named them N-1 to N-V and 3 included derived astrocytes, and we named them A-I, II and III. None of these clusters represents a single condition (Figure 1D), suggesting that the clusters were influenced but did not solely depend on the different conditions and are likely a property of the base differentiation methods. Only cluster A-III showed condition specificity, containing mostly (>90%) hiPSC-AN21. Comparing gene expression for each neuronal cluster with the rest and focusing on genes in Table 1 at FDR<0.05 with more than 1.5-fold change we found that N-I was characterized by low *NEFL, NEFM,* and *VIM,* and high *NLGN1*, N-II showed low *NEFL, NEFM, NLGN1* and *SYT1,* and high *NRXN3* and *CAMK2N1, N-III* had low *GRIA2* and *NLGN1* with high *GRIA1, NEFL, NEFM* and *VIM, N-IV* had low *GRIA2, NLGN1* and *RBFOX3* with high *ELAVL4, NEFL, NEFM, VIM* and *SLC18A3* while *N-V* showed low *NEFL* and *NEFM with high NLGN1.* When it comes to hiPSC-A cluster A-I showed low *APOE, MAP2, NEFL, PON2, SPARCL1* and *VIM* and high *GLS* and *SERPINE2*, cluster A-II showed no significant differences from the other two combined while cluster A-III showed high *APOE, NEFL, SPON2 and VIM*. All significantly different genes for each cluster are shown in suppl. Table 1.

Our observations support that iPSC-N and iPSC-A both express the appropriate marker genes suggesting similarity to in vivo excitatory neurons and astrocytes, as previous studies of these cells in bulk have shown [7,9]. However, we do find heterogeneity within hiPSC-N and hiPSC-A which is influenced but not determined by the specific conditions of differentiation that we tested here as variations of the standard differentiation protocols. This suggests that the heterogeneity is in part inherent to the differentiation protocols. Heterogeneity aside, the gene

marker data suggests a role of the presence of neurons in the maturation of astrocytes and vice-versa. Similarly, time since Ngn2-induction also seems to play an important role for the maturation of the neurons.

2. GWAS Genes expressed in iPSC-derived neurons and astrocytes

To be appropriate to use as disease models the cells created by these differentiation methods of iPSCs must express many of the genes associated with diseases involving neurons and astrocytes. We compared the list of genes showing differential expression (DE) between hiPSC-N and hiPSC-A (hiPSC-N and hiPSC-A specific genes) to those associated with neuropsychiatric illness by large genome wide association studies (GWAS) and sequencing studies for neuropsychiatric disorders. We focused on schizophrenia (SZ) and Alzheimer's disease (AD) due to the availability of large GWAS [20,21] and autism spectrum disorders (ASD) where large sequencing studies from the Simons Foundation Autism Research Initiative have identified many genes (SFARI - https://gene.sfari.org/database/human-gene/). These disorders are also within the interests of our lab and known to involve neurons and/or astrocytes. Because of the large number of genes differing in expression between the two cell types we used a stringent threshold of an adjusted p <0.001 (see methods) to focus on the genes with the highest confidence of DE, a proxy for specificity to a cell type. We also used the highest confidence genes reported for each disorder as follows: When it comes to ASD this was 207 genes with a score of 1 (highest confidence) in the SFARI database; for SZ this was 130 genes reported as genome wide significant with high confidence [20], for AD this was 38 genes at loci showing genome-wide significant association [21].

From the 28 AD-associated genes present in our dataset, 7 were among the genes significantly higher in expression in hiPSC-N and 12 with higher expression in hiPSC-A. Within genes higher in hiPSC-A this is 1.6-fold more significantly DE genes than expected but does not reach statistical significance (χ^2 - p=0.06), while there was no excess from expected in neuronal genes. Out of 106 SZ-associated genes in our dataset, 56 were among the genes expressed higher in hiPSC-N and 26 among those expressed higher in hiPSC-A, which is a highly significantly 1.55-fold more than expected for genes higher in hiPSC-N ($p=4x10^{-5}$). Finally, out of 199 ASD genes in our dataset 109 were among the genes higher in hiPSC-N and 39 among those higher in hiPSC-A, which is a 1.6-fold and strongly significant excess for hiPSC-N ($p=5x10^{-10}$) and a significant depletion (0.72-fold) for hiPSC-A (p=0.017). These results are consistent with what is currently believed for these disorders, a more important role of neurons in SZ and ASD and of astrocytes in AD. The result also supports that these hiPSC derived cells, while not equivalent to in vivo neurons and astrocytes, can be useful for modeling disease, more so than frequently used neoplastic and other cell lines. The complete set of genes with their expression in neurons and astrocytes and the comparison of the two is in suppl. Table 2

3. Differences between hiPSC-N15 and hiPSC-N21

To determine the importance of differentiation under our specific condition and its possible relevance to disease genes, we performed DE analysis between conditions and explored the properties of the genes whose expression differed significantly, starting with Differences between hiPSC-N15 and hiPSC-N21. The original protocol for generating glutamatergic neurons by Ngn2 induction reported mature neurons at 21 days post-induction [7]. Having observed

little morphological change after day 15, we decided to explore how hiPSC-N15 differ from hiPSC-N21. Shorter differentiation time not only has practical advantages, but there is a possibility that it may resemble an earlier developmental time (as supported by Table 1) perhaps more important to some diseases. The complete DE analysis results for all genes are in suppl. Table 3.

In all, 993 (7%) of 14,095 genes showed DE at an adjusted $p < 0.05$. Of those 385 were higher in hiPSC-N21 and 608 higher in hiPSC-N15. At adjusted $p < 0.1$, 571 genes were higher in hiPSC-N21 and 889 higher in hiPSC-N15. We used the longer lists of genes (adjusted $p< 0.1$) to look for enrichment of genes involved in specific biological processes using the statistical overrepresentation test of the PANTHER bioinformatics tool which allows comparisons to user-provided reference lists (in this case the list of all 14.095 genes) and corrects for multiple testing. The results for the gene lists adjusted $p < 0.05$ were similar. We used biological processes annotations from the gene ontology database (GO, http://geneontology.org/). Among the genes expressed higher in hiPSC-N15 we found significant enrichments for terms including “neurogenesis”, “neuron projection morphogenesis”, “cell morphogenesis involved in neuron differentiation”, and “regulation of neuron projection development” (see supplementary Table 4). Among the genes expressed higher in hiPSC-N21 we found among significant enrichments for the GO terms “negative regulation of neuron death”, “regulation of neurotransmitter levels”, “chemical synaptic transmission”, “neuron differentiation” and “nervous system process” (see suppl. Table 5). Our overrepresentation analysis shows that important genes in neuronal development are expressed higher in hiPSC-N15 while genes involved in neuronal function are higher in hiPSC-N21. This further supports our interpretation of the neuronal markers data shown in Table 1 in that hiPSC-N15 neurons are less mature compared to hiPSC-N21 neurons.

To gain insight on the importance of these genes in neurodevelopmental disorders we intersected them with the 130 genes reported by the Psychiatric Genomics Consortium (PGC3 data) [20] as associated with SZ with highest confidence. Of these, 105 were present in the expressed gene list and of those 18 (17%) were significantly higher (7) or lower (11) in hiPSC-N21 (Suppl. Table 6). This is significantly (1.7-fold) more DE genes than expected by chance (chi-squared $p= 0.002$). We further compared them to the list of 207 genes reported as high confidence for ASD by SFARI. Of these 199 were present in our expressed gene list and 34 (17%) were significantly higher (12) or lower (22) in hiPSC-N21 (Suppl. Table 6), which is also significantly more (1.65-fold) than expected by chance (chi-squared $p= 0.002$). This result was mostly due to a 1.75-fold genes than expected higher in hiPSC-N15 ($p = 0.006$) rather than the 1.49-fold more genes higher in hiPSC-N21 ($p = 0.149$). For the combined SZ- and ASD-associated genes the excess of observed genes was very highly significant for genes higher at hiPSC-N15 ($p=6x10^{-4}$) but still did not reach significance for genes higher in hiPSC-N21.

This excess of DE genes for ASD and SZ genes is an important observation when one uses induced neurons to study these diseases. Interestingly the strong contribution to this excess in both disorders came from genes higher in hiPSC-N15. This may suggest that, despite the artificial course of differentiation, hiPSC-N15 may resemble neurons earlier in their course to maturity and provide a better model for neurodevelopmental disorders.

4. Differences between hiPSC-N21 and hiPSC-N21A

We then explored how co-culture of **hiPSC-N** with **hiPSC-A** affects their transcriptome and whether it has a similar maturation effect as longer post-induction time. The complete transcriptome comparison results for all genes are in supplementary Table 7. Since hiPSC-N21A neurons were grown in the same plate as hiPSC-AN21 astrocytes, to avoid artifacts from any unintentional inclusion of reads from astrocytes in the hiPSC-N21A transcriptomes we excluded from this comparison genes that were expressed markedly higher in hiPSC-A than hiPSC-N at adjusted p <0.001("high hiPSC-A genes").

Overall, the expression of 237 out of 11,222 genes included in the analysis was higher in hiPSC-N21A than hiPSC-N21 at adjusted p<0.05 and 359 at adjusted p<0.1. Among the latter PANTHER bioinformatics showed a 3.3-fold enrichment for genes involved in "regulation of metal ion transport", and 1.5-fold enrichments for "cell communication", "signal transduction" and "signaling" (supplementary Table 8). The genes that were expressed significantly higher in the absence of astrocytes (hiPSC-N21) were 318 at adjusted p<0.05 and 510 at adjusted p<0.1 out of 14,857 included in the analysis. At adjusted p<0.1 there was a 7.1-fold enrichment for "central nervous system neuron axonogenesis", a 3.2 fold enrichment for "axon guidance", 2.8 fold for "axon development, 2.1 fold for neuron development and 1.6 fold for "cell differentiation" (complete results in supplementary Table 9). These results suggest that hiPSC-N21 are in an earlier stage of development than hiPSC-N21A and support the importance of the inclusion of astrocytes in maturation.

Like In the hiPSC-N15-hiPSC-N21 comparison, we compared these genes to the list of 130 high confidence SZ-associated genes reported by the PGC [20]. As opposed to the genes differing between hiPSC-N15 and hiPSC-N21, here only 5 of these genes were among those higher in hiPSC-N21A (*MSI2, CUL9, CSMD1, OPCML* and *DCC*) and 4 among those higher in hiPSC-N21 (*MAPK3, NXPH1, IL1RAPL1, GALNT17*). Similarly, when compared to the ASD genes only 8 were among those higher in hiPSC-N21A and 11 among those higher in hiPSC-N21 a both non-significantly more than expected. This might suggest that disease signal doesn't seem to come from genes that impact the functional relationship between neurons and astrocytes or could be due to reduced power.

We then tested whether the transcriptome change from hiPSC-N15 to hiPSC-N21 was similar to that from hiPSC-N21 to hiPSC-N21A. After removing "high hiPSC-A genes" (see above) there were 11,222 genes considered in both comparisons. Of these genes 281 were significantly higher in hiPSC-N21 in the first comparison and 237 were higher in hiPSC-N21A in the second comparison and 15 (6.3%) were significantly higher in both of these presumably "more mature" states. This is a highly significant overlap, 2.5-fold more than expected by chance ($p=1.4x10^{-4}$). Regarding genes expressed higher in hiPSC-N21 without astrocytes and in hiPSC-N15 (the "less mature" states) the overlap was 1.4-fold higher than expected but not statistically significant. The observed overlap for genes increasing with presumed maturity, especially since the hiPSC-N21 transcriptome is on the opposite sides of the two comparisons, suggests that the change to the induced neurons with more time post-induction bears some similarities to the changes when they are cultured with human astrocytes during their differentiation.

5. Differences between hiPSC-A0 and hiPSC-AN21

We next explored how co-culture of **hiPSC-N** with **hiPSC-A** affects the transcriptome of hiPSC-AN21 compared to hiPSC-A0. The complete transcriptome comparison results for all genes are in supplementary Table 10. Since hiPSC-AN21 were grown in the presence of hiPSC-N21A, to avoid artifacts from unintentional inclusion of reads from hiPSC-N in the hiPSC-A transcriptomes, when looking for genes higher in hiPSC-AN21 we excluded from analysis genes that were significantly higher in hiPSC-N compared to hiPSC-A (at adjusted p <0.001), the reciprocal of what we did in the HiPSC-N21/hiPSC-N21A comparison. This removed 4,675 "high hiPSC-N" genes of which 394 were also significantly higher in hiPSC-AN21 than hiPSC-A0. These included 8 SZ-GWAS genes compared to an expected 4.2 (p=0.052) possibly reflecting the general excess for genes significantly higher in hiPSC-N than hiPSC-A in SZ-GWAS, as discussed above.

In all out of 10,237 genes included in the analysis after the removal of the "high hiPSC-N" genes 411 were significantly higher in hiPSC-AN21 than hiPSC-A0 (adjusted p<0.05). At adjusted p<0.1 there were 583 genes and PANTHER bioinformatics showed multiple significant functional enrichments shown on supplementary Table 11. Notably 2.3 to 9-fold enrichments for functions like "regulation of superoxide metabolic process", "cellular oxidant detoxification", "oxidative phosphorylation", "response to oxidative stress" and "aerobic respiration" were observed, all important functions of astrocytes in the nervous system. Multiple other enrichments are not discussed here but are shown in supplementary Table 11. Due to the importance of astrocytes in AD [22] and the enrichment in astrocytic genes we observed above we also looked at the AD GWAS genes for enrichments. Twenty-one AD-associated genes were present in the reference list of genes in our comparison, and 3 of them were among the 583 significantly higher in hiPSC-AN21 (*APOE, CLU* and *CASS4)*, compared to 1.2 expected by chance. While this is not statistically significant (p=0.09) it is important to know that studying these genes using in vitro differentiated astrocytes might benefit from the inclusion of neurons in the cultures. This is particularly important for *APOE* which is a very widely studied AD gene.

When it comes to genes that were higher in hiPSC-A0 than hiPSC-AN21 out of 14,912 there were 788 at adjusted p<0.05. At adjusted p<0.1 there were 1,071 genes and PANTHER bioinformatics analysis also showed multiple significant functional enrichments, shown in supplementary Table 12. Most striking were 8 to 11-fold enrichments for immunity related genes. Of those directly relevant to neural cells, there were multiple categories involving neuron development and neural tube closure, as well as axon development and guidance.

Among 27 AD associated genes present on the set of genes tested, there were 4 AD-associated genes higher in hiPSC-A0 vs. hiPSC-AN21 (*GRN, PICALM, APH1B* and *CD2AP*). a 2.1-fold non-significant excess from expected (p=0.12), When combining genes significantly higher in hiPSC-A0 or in hiPSC-AN21 the excess from expected in GWAS genes was significant at p=0.014 highlighting that in an iPSC derived astrocyte culture the best conditions for modeling disease risk might be gene-dependent.

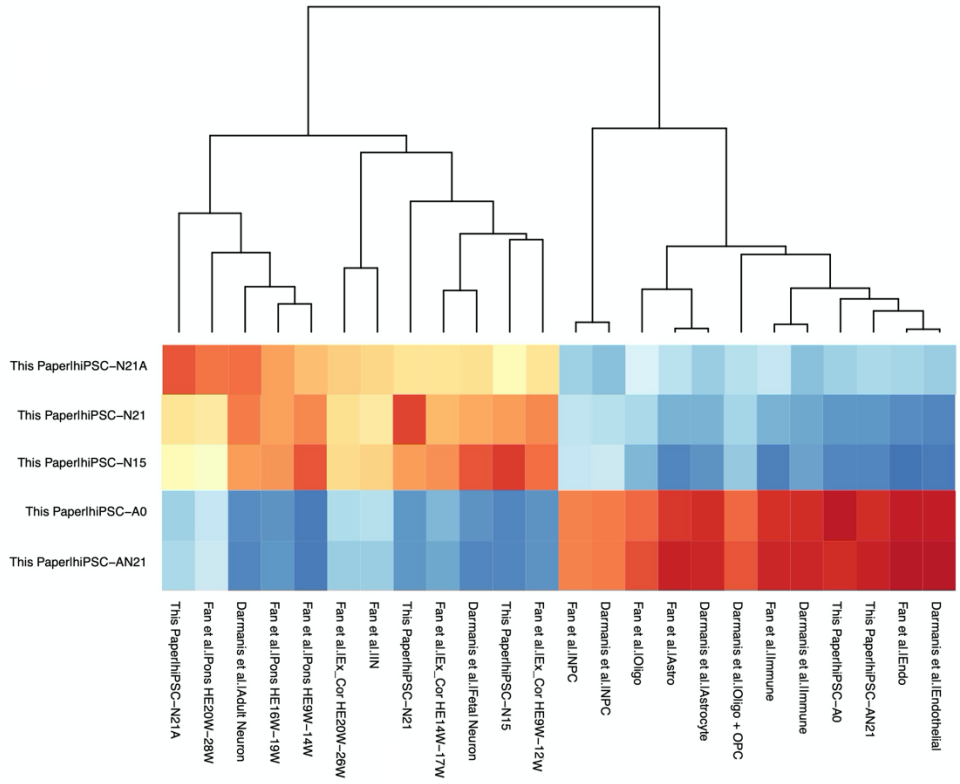

Figure 2: Nearest Neighbor analysis of our bulk and pseudobulk expression data with in vivo data from two in vivo studies. Only our 5 conditions are shown on the Y axis. Red indicates stronger correlations and blue weaker. Ex_Cor, excitatory cortical; HEW, Human embryo week; Astro, astrocytes; Oligo, oligodendrocytes; OPC oligodendrocyte precursor cells; Endo, endothelial.

6. Comparisons with in vivo datasets

Having shown that hiPSC-N and hiPSC-A express many of the expected markers for their intended cell type but also show heterogeneity within cell type we tested the resemblance of them and their subclusters to in vivo cells by comparing their transcriptomes to two previously reported in vivo datasets by Darmanis et al [23] and Fan et al [24].

Darmanis et al [23] performed single cell transcriptome analysis on adult and human fetal brain. Using unbiased grouping on 466 cells they classified them by the presence of cell-type specific markers in one of the following categories: Oligodendrocyte precursor cells (OPCs), oligodendrocytes, astrocytes, microglia, neurons, endothelial cells, replicating neuronal progenitors, and quiescent newly born neurons. Two remaining unbiased clusters contained genes characteristic of more than one distinct cell type and we did not include them in our analysis. Fan et al [23] performed single-cell transcriptome profiling 12,541 cells from the four cortical lobes and pons during human embryonic and fetal development from the 7th to the 28th gestational week (GW), similarly classified to neural subtypes. We first performed a meta neighbor analysis [25] across the three datasets to explore similarities across cell types in the different experiments. The complete heatmap is shown in supplementary Figure 1, while Figure 2 shows part of it, restricting the x-axis to our 5 conditions. As expected, the cells differentiated towards neuronal types are closest to the in vivo neuronal cell types, including both cortical and pontine. hiPSC-N21A are closest to Darmanis et al adult neurons but also close to Fan et al late gestation pontine neurons (Figure 2). Interestingly the Darmanis et al adult neurons are also

close to Fan et al pontine neurons (supplementary Figure 1). Our hiPSC-N21 are also closest to Darmanis et al adult neurons, with the next closest neighbors being Fan et al 9-12 GW excitatory cortical neurons and GW 9-14 pontine neurons. Our hiPSC-N15 neurons are closest to Darmanis et al fetal neurons and Fan et al early pontine neurons with the next best neighbor being Fan et al early cortical neurons. This is of particular interest as it suggests that hiPSC-N15 to some extent recapitulate an earlier developmental state despite the forced differentiation by Ngn2. It is particularly interesting in view of our observation that genes expressed higher at hiPSC-N15 contain the strongest excess for genes identified for neurodevelopmental disorders such as schizophrenia and autism (see section 3). The hiPSC-A, both hiPSC-A0 and hiPSC-AN21 were similar to the astrocytes of both Darmanis et al and Fan et al but were also similar to the endothelial and immune cells of both studies and a little less so to oligodendrocytes and NPCs (Figure 2). While this was a surprising result, the astrocytes from both in vivo studies also showed strong similarities to these cell types (suppl. Figure 1, suggesting this might not be an irregularity of our hiPSC-A0 and hiPSC-AN21 but rather a property of these cell types in vivo as well. The hiPSC-AN21 showed somewhat stronger similarities to most of the above in vivo cell types including astrocytes.

To further dissect these relationships, we performed Seurat CCA-based integration analysis on our cells along with cells from the Fan study. This integration analysis was carried out separately within the neuronal and then non-neuronal cell types (as delineated by Figure 2) to focus on specific lineage relationships. In the neuronal analysis (Figure 3), separation across the first UMAP dimension placed our neurons between excitatory cortical and PRPH-expressing pontine neurons in vivo, supporting their excitatory neuronal identity but indicating a deviation from bona fide in vivo cortical neuron identity. This is consistent with recent studies of scRNA-seq data in Ngn2-indced neurons which concluded that Ngn2-induction produces PRPH-expressing sensory neurons [26,27]. Interestingly, the second UMAP dimension aligned with increasing maturity for both in vivo pontine and cortical neurons, while in our cells followed the sequence of hiPSC-N15→hiPSC-N21→hiPSC-N21A. This further supports our conclusion that longer time from induction and culture with astrocytes promotes neuronal maturation.

In the non-neuronal integration analysis of our hiPSC-A data with data from Fan et al (Figure 4) there were multiple clusters. The hiPSC-A0 , and hiPSC-AN21 cells that were in Louvain clusters A-I and A-II clustered near each other and were surrounded by three clusters of vascular and endothelial smooth muscle cells. Louvain cluster A-III which was mostly comprised of hiPSC-AN21 cells (Figure 1D) was adjacent to the cluster of in vivo Astrocytes, suggesting that many, though not all, hiPSC-AN21 achieve higher resemblance to in vivo astrocytes likely due to the coculture with induced neurons. In vivo oligodendrocytes and immune cells were in separate clusters with many distinct sub clusters.

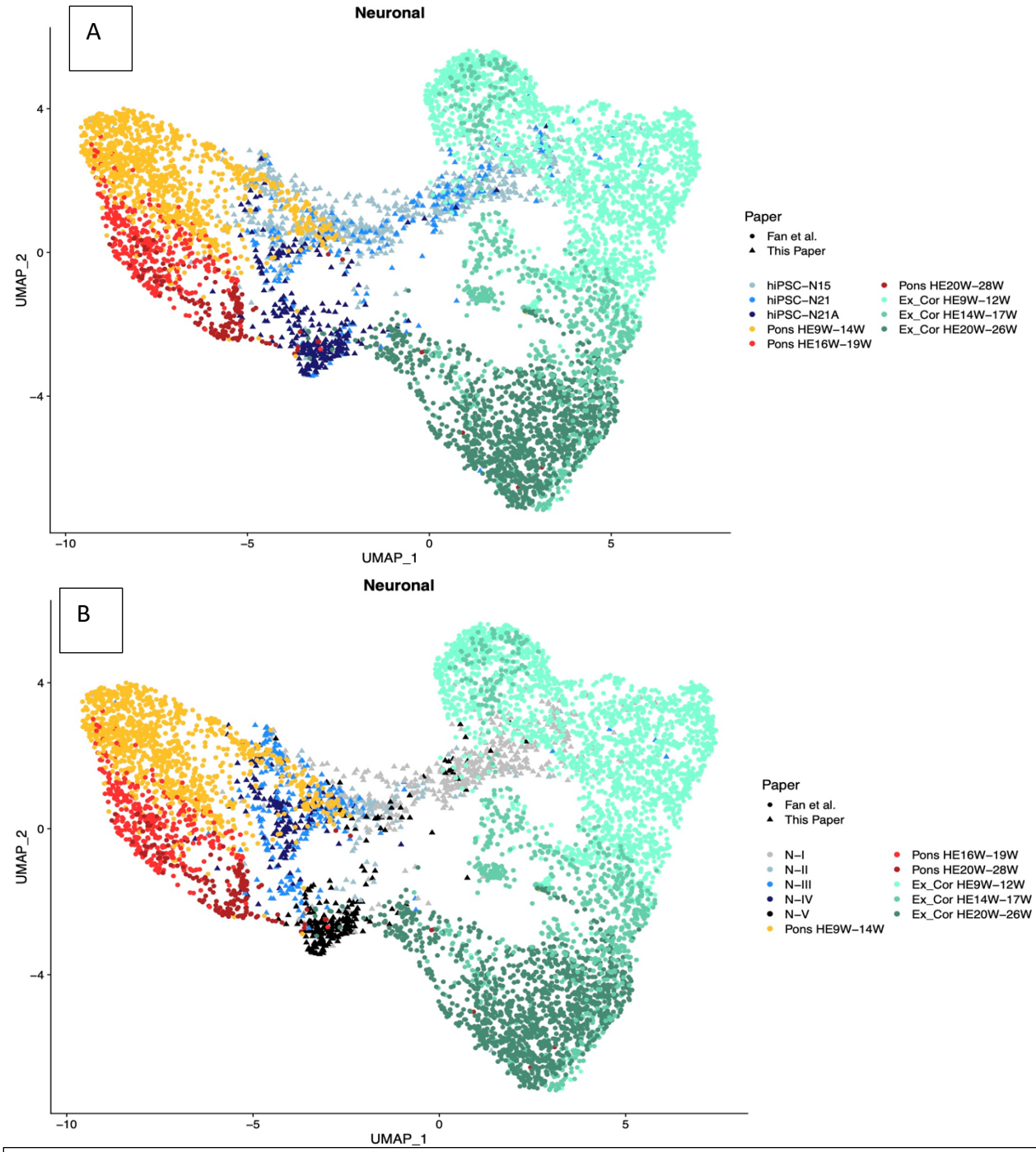

Figure 3: Seurat integration analysis of our cells together with two in vivo datasets. For simplicity only in vivo and iPSC derived neuronal cells are included. (A) our cells colored by condition. (B) our cells colored by Louvain cluster. Ex_Cor, excitatory cortical; HEW, Human embryo week.

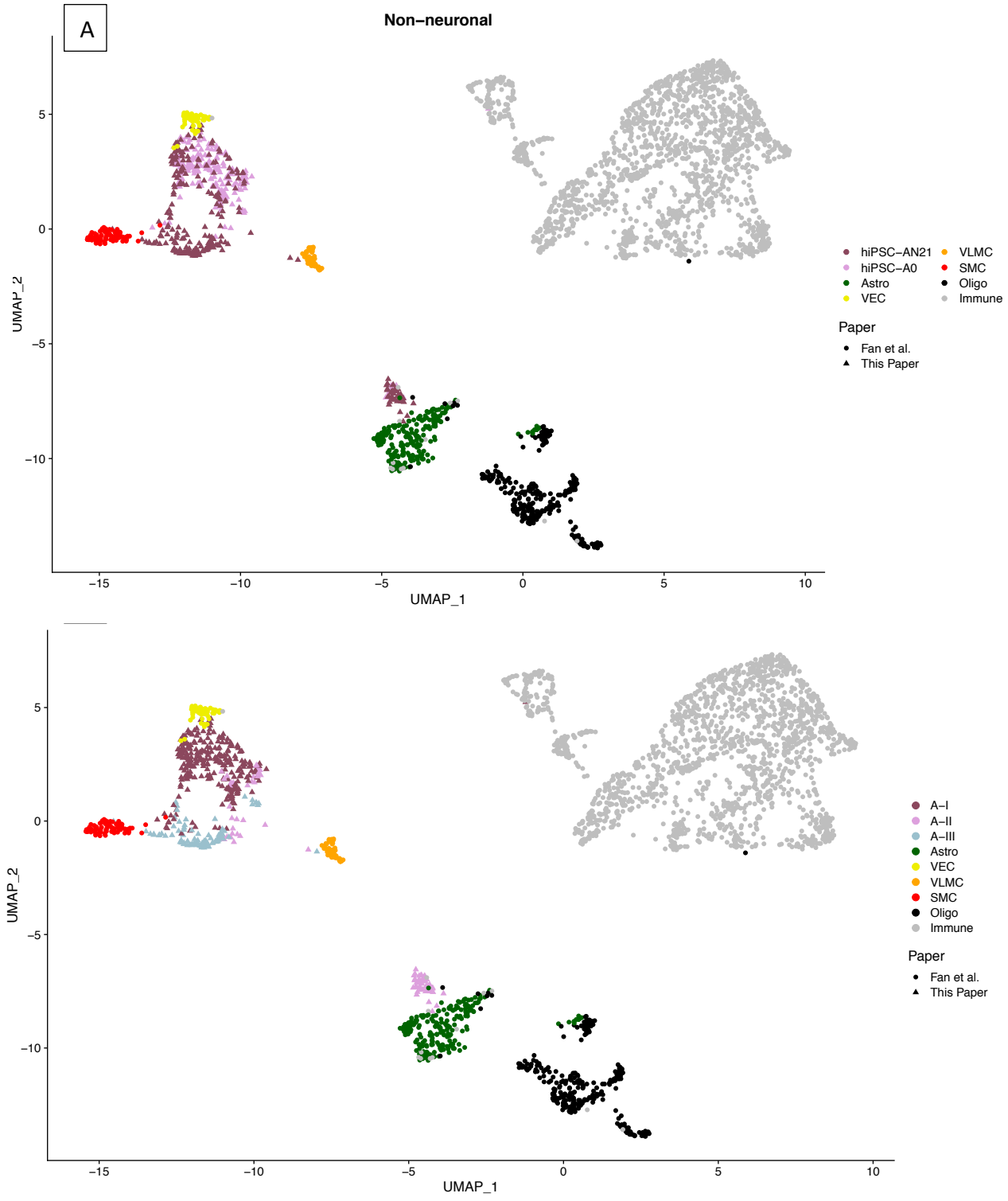

Figure 4: Seurat integration analysis of our hiPSC-A with two in vivo datasets. Only in vivo non neuronal cells are included. (A) our cells colored by condition. (B) our cells colored by Louvain cluster. SMC, smooth muscle cells; VEC, vascular endothelial cells; VLMC, vascular leptomeningeal cells. Astro, astrocytes; Oligo, oligodendrocytes

Together with recently published scRNA-seq data in the NGN2-induction system [26,27], these results indicate that while Ngn2-induced neurons are excitatory neurons, they share transcriptional elements with multiple in vivo neuronal lineages. To explore the reproducibility of this complex induced neuronal phenotype, we performed a third integration analysis bringing together our Ngn2-induced cell data with this recent scRNA-seq data (Figure 5). The integrated UMAP highlights in red different cells from each study and recapitulates elements of the heterogeneity we observe in our hiPSC-N in the other studies and shows that they co-localize across a temporal progression of cellular maturity. Hence, although the Ngn2-induced neuronal state appears to include a complex combination of in vivo transcriptional programs and is reproducible across individual induction experiments and different laboratories.

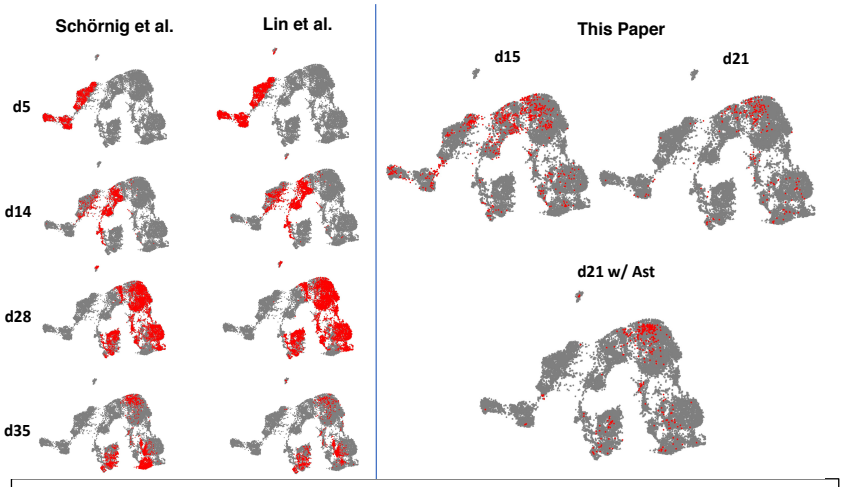

Figure 5: Integration of our data with other Ngn2-induced single cell data. The complete map is in gray and Subsets from each study are highlighted in red

DISCUSSION

Our single cell analysis across the 4 different conditions showed that hiPSC-N and hiPSC-A could be sharply separated based on their transcriptome (Figure 1) and expressed appropriate cell type markers (Figure 2). Both hiPSC-N and hiPSC-A however showed heterogeneity and could be divided to sub-clusters that were influenced but not determined by the different conditions we tested. Ngn2-induced neurons do not represent a singular in vivo neuronal cell type, but rather express transcriptional characteristics of both cortical excitatory neurons and more posterior (pons) neuronal fates, as well as markers of other diverse neuronal subtypes. While this is consistent with recent in-depth single-cell analyses that concluded Ngn2-induced neurons take on specific sensory fates [16], we find that Ngn2-induction produces a type of neuronal cell that includes transcriptional elements of multiple in vivo neuronal cell types. So, while their neuronal identity makes them a better model for studying brain disorders than previously used neoplastic or immortalized cell lines, one still needs to be cautious in reaching conclusions and further work would be helpful in refining differentiation protocols. When it comes to iPSC-Ast, we find that they take on a transcriptional state resembling several in vivo cell types (including endothelial cells, immune cells and astrocytes). These cell types however also appear to have similarities in the vivo datasets (suppl. Figure 1), so this should not be seen as a big deviation from astrocytic identity. Co-culture with neurons pushes a subpopulation of the in hiPCS-A toward a state more similar to bona fide in vivo astrocytes, however even with

co-culture only some of the cells take on a clear astrocytic identity, so caution is necessary when studying them in bulk.

Unique aspects of our study include the examination of two post Ngn2 induction time points, the co-culture of hiPSC-N with hiPSC-A and the integration with in vivo datasets. This allowed us to show that both the longer post-induction culture time of the Ngn2-induced neurons, and the inclusion of hiPSC-A contributed to expression profiles closer to mature neurons. This was observed not only in terms of expression of neuronal markers (Table 1), but also in terms of the functions of the DE genes and of similarity to in-vivo neurons at different maturity states (Figures 2,3). We made the same observation for hiPSC-A, where the co-culture with hiPSC-N21 also appeared to increase their maturity, and hiPSC-AN21 were the only ones that distinctly co-clustered with human in vivo astrocytes. The difference between conditions across most astrocytic marker-gene expression was striking, more so than in hiPSC-N, (Figure 2) and it included *APOE*, a very important astrocytic gene in the study of AD. In contrast to this observation though, the excess of genes associated with AD was observed among those expressed lower, not higher, in hiPSC-AN21 compared to hiPSC-A0. It is therefore unclear whether one should prefer the co-cultured neurons for modeling AD, a decision that should probably be made on a gene and experiment specific basis.

Our comparisons with in vivo datasets showed that the NgN2 induced neurons are similar to in vivo excitatory neurons, with proximity to both cortical and pontine in vivo neurons, validating them as a cellular model. We find it particularly interesting that the trajectory from hiPSC-N15 to hiPSC-N21 to hiPSC-N21A was along the same axis with the pontine and cortical neurons development during fetal life (Figure 3), with the hiPSC-N15 being closer to fetal than adult neurons, which has significant implications for the use of induced neurons for the study of disease, as we will discuss below. The comparison of the hiPSC-A0 and hiPSC-AN21 with in vivo datasets confirmed their similarity to in vivo astrocytes but additionally showed strong similarities to immune and endothelial cells. While these could be interpreted as deviations from the true cell type, we also found the same similarities within the in vivo datasets, i.e., the astrocytes in both in vivo datasets were also similar to the immune and endothelial cells. It appears that this maybe an inherent property of these cell types and not a problem in the differentiation process. The overall profile of expression of astrocytic marker genes (Table 1) and the excess (albeit only suggestive) of AD-associated genes further supports this interpretation.

Our search for overrepresentation in GWAS genes was triggered by the main goal of this paper, to explore the use of an in vitro differentiation system for the study of psychiatric disorders, specifically neurodevelopmental (SZ, ASD) and neurodegenerative (AD) diseases. When it comes to neuron versus astrocyte-predominant genes the results were unequivocal. Genes expressed higher in neurons contained an excess SZ- and ASD-associated genes validating them as a platform to model these disorders, while a suggestive excess (which could be due to lack of power) was seen in astrocytic genes for AD-associations, and interestingly this was significant for genes expressed higher in hiPSC-A0 than hiPSC-AN21. Regarding finer distinctions based on days post-induction and co-culture we found that genes expressed higher in hiPSC-N15 than hiPSC-N21 showed the highest content of neurodevelopmental disorder genes. The induction by Ngn2 is far from the normal course of differentiation and maturation of

neurons in vivo, yet this along with the similarity of hiPSC-N15 to fetal neurons suggests that between day 15 and 21 the course may be similar and hiPSC-N15 maybe a better time to model for SZ and ASD than hiPSC-N21.

We have harnessed the power of single cell sequencing and iPS cell differentiation to successfully gain important insights for disease modeling. While it will take a lot more studies and time for our results to be validated through additional experiments and actual disease modeling, we anticipate that this study will be an important additional guide for navigating the complexities of complex brain disorders and improving differentiation protocols to achieve the optimal disease models.

MATERIALS AND METHODS

Induced pluripotent stem cell (iPSC) culture and maintenance

Human BC1, an iPSC line which obtained from Dr. Linzhao Cheng's lab at Johns Hopkins School of Medicine, was used in the study. This is an established cell line with published results. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5746148/) (Cai etal 2018). Cells were cultured in StemFlex media (Gibco) on 6-well tissue culture plates coated with Laminin (Biolamina). Cells were dissociated with StemPro Accutase (Gibco) into single cell suspension and seeded in required density for the experiment (see nelow). The Rock inhibitor Y-27632 dihydrochloride (Tocris) was added on the first day of passage at a concentration of 10uM. Cultured cells were tested to ensure they lack mycoplasma contamination.

Ngn2 lentivirus transduction

Ngn2 and rTTA virus were procured from the University of Pennsylvania Core store. This virus has been previously reported to be successfully used for induced neural differentiation by our (Avramopoulos) laboratory [28]. It was initially reported by the Sudhof laboratory [7] who first discovered that forced expression of this single transcription factor Ngn2 can convert iPSCs into functional neurons with very high yield in 21 days. 250,000 BC1 iPSC cells were plated in each well of a 6 well plate and grown in Stem Flex media supplemented with Rock inhibitor. Ngn2 lentiviral infection using Polybrene (Santa Cruz) was done 24 hours post seeding. Briefly cells were fed with 2ml of fresh media and 2ul of Polybrene (1ug/ml) stock was added per well. To attain a MOI of 1-10 different volumes of both the Ngn2 and rTTA virus was added per well. In 4 of the 6 wells, leaving one as a polybrene-only control following amounts of EACH virus was added: 3ul, 5ul and 10ul. The virus infected cells were expanded and frozen stocks made for future differentiation. We selected for the 10ul transduced Ngn2-BC1 cells for optimal neural differentiation.

Neuronal differentiation of Ngn2 transduced iPSCs

250,000 Ngn2 transduced BC1 cells were plated on laminin coated 6 well plates **(DIV -2)**. Cells were fed with fresh Stem Flex media the next day **(DIV -1)**. Ngn2 expression was induced by Doxycycline on DIV 0 using an induction media consisting of DMEM/F12 (Thermo Fisher), N2 (Thermo Fisher), D-Glucose (Thermo Fisher), 2-βME (Life technologies), Primocin (Invivogen), BDNF (10ng/ml, Peprotech), NT3 (10ng/ml, Peprotech), Laminin (200ng/ml, Millipore Sigma) and Doxycycline (2ug/ml, Sigma). A

puromycin selection was done on these cells on DIV 1, 24 hours post Doxycycline induction using the same induction media supplemented with puromycin (5ug/ml). Surviving cells were harvested on DIV 2 and plated on matrigel coated 24 well plates at a concentration of 100,000 cells/well in neural differentiation media consisting of Neurobasal media (Thermo Fisher), B27 (Thermo Fisher), Glutamax (Thermo Fisher), Penn/Strep (Thermo Fisher), D-Glucose (Thermo Fisher), BDNF (10ng/ml), NT3 (10ng/ml), Laminin (200ng/ml) and Doxycycline (2ug/ml). Cells were fed with a 50% media exchange of neural differentiation media every other day till DIV 12. Cells were treated with 2uM Cytosineβ-D-arabinofuranoside hydrochloride (Ara-C) on DIV 4 to arrest proliferation and eliminate non neuronal cells in the culture. Doxycycline induction was initiated at DIV 0 and continued till DIV 12 after which it was discontinued and cells were fed every two days thereafter till DIV 21 with Neural maturation media consisting of Neurobasal media A (Thermo Fisher), B27, Glutamax, Penn/Strep, Glucose Pyruvate mix (1:100, final conc of 5mM glucose and 10mM sodium pyruvate), BDNF (10ng/ml), NT3 (10ng/ml) and Laminin (200ng/ml). Neurons were harvested by DIV 15 or 21. Four conditions were set up for this experiment which are as follows: i) Neurons only (DIV 21) ii) Neurons only (DIV 15) iii) Astrocytes only iv) Neurons and astrocyte coculture. Astrocytes were added on top of the differentiating cells on DIV 5 at a concentration of 50,000 cells/well in the coculture experiment. 2uM Ara-C treatment was repeated on DIV 7 for the coculture experiment and media changed every other day thereafter till DIV 21. For the astrocyte only condition astrocytes were seeded at 50,000 cells per well of a 24 well plate and fed with Neural differentiation media and allowed to grow till 80% confluent before adding 2uM AraC. Media was changed every other day till DIV 21 when neurons are ready to harvest. Neurons were collected using Accutase and passed through a cell strainer and counted to receive the optimal number of cells.

Neural differentiation of hiPSCs via embryoid body (EB) formation

Neural differentiation of embryoid bodies (EBs) was performed as previously described [29] with modifications. Briefly, EB formation was performed by the forced aggregation method. To this goal, PSC lines were cultured in feeder-free conditions as monolayers with E8 medium and passaged every 3 days with TrypLE. For the production of uniform-size EBs, iPSCs grown for 3–10 passages were counted and seeded at 5,000 cells per well in 96-well, V-bottom uncoated plates (249952; NUNC, Rochester, NY). For induction of neural differentiation, EBs were grown in suspension for 7-8 days followed by adherence to Matrigel-coated plates in the Neural Induction Medium (NIM) consisting of DMEM/F12 (GIBCO, 11320033), 2 mM l-glutamine, 0.1% bovine serum albumin (Fraction V; Sigma-Aldrich), 1% NEAA, 2% B27 without retinoic acid (GIBCO), 1% N2 supplement (GIBCO), LDN193189 (Peprotech) throughout culture, and 10 μM SB431542 (Tocris Bioscience, Bristol, United Kingdom). Numerous rosette structures were formed 2-3 days after the adherent culture of EBs.

Isolation and culture of neural precursor cells

Neural rosettes were manually collected with stretched glass Pasteur pipettes and expanded as monolayer cultures of neural precursors (NPCs). Briefly, EB-derived neural rosettes were dissociated into single cells with Accutase for 5 min at 37°C and plated on Matrigel or polyornithine/laminin-coated plates in the NIM complete medium supplemented with FGF2 (10 ng/mL) and epidermal growth factor (EGF) (10 ng/mL; PeproTech, Rocky Hill, NJ). Cells were expanded for several passages as a homogeneous population of NPCs.

Astrocytic differentiation

Human BC1iPSC line was differentiated into astrocytes as previously described [9]. Briefly, NPCs dissociated to single cells were seeded at 15,000 cells/cm^2 density on Matrigel coated plate in complete astrocytic differentiation medium (ScienCell Research Laboratories cat. No 1801), astrocyte medium (ScienCell Research Laboratories cat. No 1801-b), 2% fetal bovine serum (ScienCell Research Laboratories cat. No 0010), astrocyte growth supplement (ScienCell Research Laboratories cat. No 1852). The cells passaged in this density for the first 30 days and fed every other day. Following this period, the astrocytes could be passaged in a 1:3 ratio and expanded for up to 120 days in the same medium.

Single cell sequencing

6 wells in a 24-well plates of Neurons were grown for 15 days post Ngn2 induction, 6 wells in a 24well plates of Neurons were grown for 21 days post Ngn2 induction and 6 wells in a 24-well plates of Neurons were grown for 21 days post Ngn2 induction with the addition of astrocytes on day5 at a density of 50,000 cells/well After dissociation with Accutase, single cell suspensions for 10x libraries were loaded onto the 10x Genomics Chromium Single Cell system using the v2 chemistry per manufacturer's instructions [30,31]. Estimations of cellular concentration and live cells in suspension was made through Trypan Blue staining and use of the Countess II cell counter (ThermoFisher). Dissociated single iPSCs were passed through a 40um filter and used as input for the 10x chromium v2 3' gene expression kit (10x genomics), targeting 1,000 cells per sample. Libraries were prepared according to the manufacturer's instructions and uniquely indexed. Libraries were quantified on the Nanodrop platform and sized using the Agilent 2100 Bioanalyzer RNA nano system. Barcoded libraries were pooled and sequenced on an S1 flowcells on a NovaSeq 6000 (Illumina) to an average depth of ~1.33x10^8 (+- 3.92x10^7) paired-end reads per sample. Raw reads were pseudoaligned to the Gencode reference human transcriptome (v31; www.gencodegenes.org/human/) using kallisto (default parameters plus -t 4) and collapsed to individual UMIs using bustools correct (default parameters plus -t 4; 10x v2 whitelist) and bustools count (default parameters plus -t 4). Cells were filtered from empty droplets using estimated knee plot inflection point UMI cutoffs (DropletUtils) with the minimum UMI thresholds ranging between 1608-7616 across samples. BUS records from each sample were aggregated to a unified counts Table, used as input for the monocle3 R/Bioconductor single cell framework (https://cole-trapnell-lab.github.io/monocle3/), and processed using default workflow settings.

Differential gene expression analysis in scRNA-seq data

Differential gene expression analysis across cell types in our scRNA-seq data was performed using DESeq2 [32] along with specific recommendations for its application to scRNA-seq data using additional methods in the zinbwave [33] and scran [34] packages at: http://bioconductor.org/packages/devel/bioc/vignettes/DESeq2/inst/doc/DESeq2.html#recommendations-for-single-cell-analysis, which draws on analyses and conclusion from Van den Berge, Zhu et al, and Ahlmann & Huber [35-37]. Briefly, the computeSumFactors() function in the scran package was used to calculate size factors that were passed to the zinbwave() function

and then output was passed onto the DESeq2 functions DESeqDataSet() and DESeq(). The DESeq() differential gene expression function was implemented using test="LRT" rather than the Wald test for significance testing, along with these scRNAseq-specific argument values: useT=TRUE, minmu=1e-6, and minReplicatesForReplace=Inf.

Single-cell clustering and dimensionality reduction

For the original scRNA-seq data from this report, UMI count matrices were processed and analyzed using the Seurat package (v4.1.0) in R [38]. A total of 1631 cells were included in this analysis. After datasets were normalized, the top 10,000 variable genes were selected for further analysis using the variance stabilizing transformation (vst) method. 2D visualization of our data was accomplished using principal components analysis (PCA) followed by Uniform Manifold Approximation and Projection (UMAP) of these PCs.

Seurat integration analysis

Integration of our scRNa-seq data with public scRNA-seq UMI count data[24] (GSE120046) was carried out using canonical correlation analysis (CCA) in Seurat [39,40]. Integrated datasets were scaled and PCA was performed. We chose the first 10 PC's for use in non-linear dimensionality reduction by identifying the elbow on a Skree plot of the first 30PCs. 2D visualization of the integrated data was accomplished using the Uniform Manifold Approximation and Projection (UMAP) algorithm on these 10 PCs.

MetaNeighbor Analysis

Cell-type replicability analysis across datasets was performed using the MetaNeighbor (v1.1.0; Crow et al 2018 [25]) package in R. We used unsupervised MetaNeighbor to first determine intersecting highly variable genes across datasets, and then used the Spearman correlation network as described in Crow et al [25] to determine replicability.

ACKNOWLEDGEMENTS: This work was supported by NIMH grants R01 MH113215 and RF1 MH122936 to DA. Data sharing and visualization via NeMO Analytics was supported by grants R24MH114815 and R01DC019370

REFERENCES

1. Das, D.; Feuer, K.; Wahbeh, M.; Avramopoulos, D. Modeling Psychiatric Disorder Biology with Stem Cells. *Curr Psychiatry Rep* **2020**, *22*, 24, doi:10.1007/s11920-020-01148-1.
2. Hartley, B.J.; Tran, N.; Ladran, I.; Reggio, K.; Brennand, K.J. Dopaminergic differentiation of schizophrenia hiPSCs. *Mol Psychiatry* **2015**, *20*, 549-550, doi:10.1038/mp.2014.194.
3. Kerr, C.L.; Letzen, B.S.; Hill, C.M.; Agrawal, G.; Thakor, N.V.; Sterneckert, J.L.; Gearhart, J.D.; All, A.H. Efficient differentiation of human embryonic stem cells into oligodendrocyte progenitors for application in a rat contusion model of spinal cord injury. *Int J Neurosci* **2010**, *120*, 305-313, doi:10.3109/00207450903585290.

4. Krencik, R.; Zhang, S.C. Directed differentiation of functional astroglial subtypes from human pluripotent stem cells. *Nat Protoc* **2011**, *6*, 1710-1717, doi:10.1038/nprot.2011.405.
5. Liu, Y.; Liu, H.; Sauvey, C.; Yao, L.; Zarnowska, E.D.; Zhang, S.C. Directed differentiation of forebrain GABA interneurons from human pluripotent stem cells. *Nat Protoc* **2013**, *8*, 1670-1679, doi:10.1038/nprot.2013.106.
6. Pandya, H.; Shen, M.J.; Ichikawa, D.M.; Sedlock, A.B.; Choi, Y.; Johnson, K.R.; Kim, G.; Brown, M.A.; Elkahloun, A.G.; Maric, D.; et al. Differentiation of human and murine induced pluripotent stem cells to microglia-like cells. *Nat Neurosci* **2017**, *20*, 753-759, doi:10.1038/nn.4534.
7. Zhang, Y.; Pak, C.; Han, Y.; Ahlenius, H.; Zhang, Z.; Chanda, S.; Marro, S.; Patzke, C.; Acuna, C.; Covy, J.; et al. Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* **2013**, *78*, 785-798, doi:10.1016/j.neuron.2013.05.029.
8. Simone, N.L.; Bonner, R.F.; Gillespie, J.W.; Emmert-Buck, M.R.; Liotta, L.A. Laser-capture microdissection: opening the microscopic frontier to molecular analysis. *Trends Genet* **1998**, *14*, 272-276, doi:10.1016/s0168-9525(98)01489-9.
9. Tcw, J.; Wang, M.; Pimenova, A.A.; Bowles, K.R.; Hartley, B.J.; Lacin, E.; Machlovi, S.I.; Abdelaal, R.; Karch, C.M.; Phatnani, H.; et al. An Efficient Platform for Astrocyte Differentiation from Human Induced Pluripotent Stem Cells. *Stem Cell Reports* **2017**, *9*, 600-614, doi:10.1016/j.stemcr.2017.06.018.
10. Orvis, J.; Gottfried, B.; Kancherla, J.; Adkins, R.S.; Song, Y.; Dror, A.A.; Olley, D.; Rose, K.; Chrysostomou, E.; Kelly, M.C.; et al. gEAR: Gene Expression Analysis Resource portal for community-driven, multi-omic data exploration. *Nat Methods* **2021**, *18*, 843-844, doi:10.1038/s41592-021-01200-9.
11. Cardona, C.; Sanchez-Mejias, E.; Davila, J.C.; Martin-Rufian, M.; Campos-Sandoval, J.A.; Vitorica, J.; Alonso, F.J.; Mates, J.M.; Segura, J.A.; Norenberg, M.D.; et al. Expression of Gls and Gls2 glutaminase isoforms in astrocytes. *Glia* **2015**, *63*, 365-382, doi:10.1002/glia.22758.
12. Lee, M.; Schwab, C.; McGeer, P.L. Astrocytes are GABAergic cells that modulate microglial activity. *Glia* **2011**, *59*, 152-165, doi:10.1002/glia.21087.
13. Kulik, A.; Vida, I.; Lujan, R.; Haas, C.A.; Lopez-Bendito, G.; Shigemoto, R.; Frotscher, M. Subcellular localization of metabotropic GABA(B) receptor subunits GABA(B1a/b) and GABA(B2) in the rat hippocampus. *J Neurosci* **2003**, *23*, 11026-11035.
14. Verrier, J.D.; Jackson, T.C.; Gillespie, D.G.; Janesko-Feldman, K.; Bansal, R.; Goebbels, S.; Nave, K.A.; Kochanek, P.M.; Jackson, E.K. Role of CNPase in the oligodendrocytic extracellular 2',3'-cAMP-adenosine pathway. *Glia* **2013**, *61*, 1595-1606, doi:10.1002/glia.22523.
15. Wislet-Gendebien, S.; Wautier, F.; Leprince, P.; Rogister, B. Astrocytic and neuronal fate of mesenchymal stem cells expressing nestin. *Brain Res Bull* **2005**, *68*, 95-102, doi:10.1016/j.brainresbull.2005.08.016.
16. Cho, J.M.; Shin, Y.J.; Park, J.M.; Kim, J.; Lee, M.Y. Characterization of nestin expression in astrocytes in the rat hippocampal CA1 region following transient forebrain ischemia. *Anat Cell Biol* **2013**, *46*, 131-140, doi:10.5115/acb.2013.46.2.131.

17. Britanova, O.; de Juan Romero, C.; Cheung, A.; Kwan, K.Y.; Schwark, M.; Gyorgy, A.; Vogel, T.; Akopov, S.; Mitkovski, M.; Agoston, D.; et al. Satb2 is a postmitotic determinant for upper-layer neuron specification in the neocortex. *Neuron* **2008**, *57*, 378-392, doi:10.1016/j.neuron.2007.12.028.
18. Vallano, M.L.; Beaman-Hall, C.M.; Mathur, A.; Chen, Q. Astrocytes express specific variants of CaM KII delta and gamma, but not alpha and beta, that determine their cellular localizations. *Glia* **2000**, *30*, 154-164, doi:10.1002/(sici)1098-1136(200004)30:2<154::aid-glia5>3.0.co;2-s.
19. Blondel, V.D.; Guillaume, J.-L.; Lambiotte, R.; Lefebvre, E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* **2008**, *2008*, P10008, doi:10.1088/1742-5468/2008/10/p10008.
20. Ripke, S.; Walters, J.T.R.; O'Donovan, M.C.; Consortium, T.S.W.G.o.t.P.G. Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. *bmedRxiv* **2020**, *doi: https://doi.org/10.1101/2020.09.12.20192922*, doi:10.1101/2020.09.12.20192922.
21. Wightman, D.P.; Jansen, I.E.; Savage, J.E.; Shadrin, A.A.; Bahrami, S.; Holland, D.; Rongve, A.; Borte, S.; Winsvold, B.S.; Drange, O.K.; et al. A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat Genet* **2021**, *53*, 1276-1282, doi:10.1038/s41588-021-00921-z.
22. Monterey, M.D.; Wei, H.; Wu, X.; Wu, J.Q. The Many Faces of Astrocytes in Alzheimer's Disease. *Front Neurol* **2021**, *12*, 619626, doi:10.3389/fneur.2021.619626.
23. Darmanis, S.; Sloan, S.A.; Zhang, Y.; Enge, M.; Caneda, C.; Shuer, L.M.; Hayden Gephart, M.G.; Barres, B.A.; Quake, S.R. A survey of human brain transcriptome diversity at the single cell level. *Proc Natl Acad Sci U S A* **2015**, *112*, 7285-7290, doi:10.1073/pnas.1507125112.
24. Fan, X.; Fu, Y.; Zhou, X.; Sun, L.; Yang, M.; Wang, M.; Chen, R.; Wu, Q.; Yong, J.; Dong, J.; et al. Single-cell transcriptome analysis reveals cell lineage specification in temporal-spatial patterns in human cortical development. *Sci Adv* **2020**, *6*, eaaz2978, doi:10.1126/sciadv.aaz2978.
25. Crow, M.; Paul, A.; Ballouz, S.; Huang, Z.J.; Gillis, J. Characterizing the replicability of cell types defined by single cell RNA-sequencing data using MetaNeighbor. *Nat Commun* **2018**, *9*, 884, doi:10.1038/s41467-018-03282-0.
26. Lin, H.C.; He, Z.; Ebert, S.; Schornig, M.; Santel, M.; Nikolova, M.T.; Weigert, A.; Hevers, W.; Kasri, N.N.; Taverna, E.; et al. NGN2 induces diverse neuron types from human pluripotency. *Stem Cell Reports* **2021**, *16*, 2118-2127, doi:10.1016/j.stemcr.2021.07.006.
27. Schornig, M.; Ju, X.; Fast, L.; Ebert, S.; Weigert, A.; Kanton, S.; Schaffer, T.; Nadif Kasri, N.; Treutlein, B.; Peter, B.M.; et al. Comparison of induced neurons reveals slower structural and functional maturation in humans than in apes. *Elife* **2021**, *10*, doi:10.7554/eLife.59323.
28. Das, D.; Peng, X.; Lam, A.N.; Bader, J.S.; Avramopoulos, D. Transcriptome analysis of human induced excitatory neurons supports a strong effect of clozapine on cholesterol biosynthesis. *Schizophr Res* **2021**, *228*, 324-326, doi:10.1016/j.schres.2020.12.041.
29. Mahairaki, V.; Ryu, J.; Peters, A.; Chang, Q.; Li, T.; Park, T.S.; Burridge, P.W.; Talbot, C.C., Jr.; Asnaghi, L.; Martin, L.J.; et al. Induced pluripotent stem cells from familial

Alzheimer's disease patients differentiate into mature neurons with amyloidogenic properties. *Stem Cells Dev* **2014**, *23*, 2996-3010, doi:10.1089/scd.2013.0511.

30. Zheng, G.X.; Terry, J.M.; Belgrader, P.; Ryvkin, P.; Bent, Z.W.; Wilson, R.; Ziraldo, S.B.; Wheeler, T.D.; McDermott, G.P.; Zhu, J.; et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **2017**, *8*, 14049, doi:10.1038/ncomms14049.
31. Clark, B.S.; Stein-O'Brien, G.L.; Shiau, F.; Cannon, G.H.; Davis-Marcisak, E.; Sherman, T.; Santiago, C.P.; Hoang, T.V.; Rajaii, F.; James-Esposito, R.E.; et al. Single-Cell RNA-Seq Analysis of Retinal Development Identifies NFI Factors as Regulating Mitotic Exit and Late-Born Cell Specification. *Neuron* **2019**, *102*, 1111-1126 e1115, doi:10.1016/j.neuron.2019.04.010.
32. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **2014**, *15*, 550, doi:10.1186/s13059-014-0550-8.
33. Risso, D.; Perraudeau, F.; Gribkova, S.; Dudoit, S.; Vert, J.P. A general and flexible method for signal extraction from single-cell RNA-seq data. *Nat Commun* **2018**, *9*, 284, doi:10.1038/s41467-017-02554-5.
34. Lun, A.T.; McCarthy, D.J.; Marioni, J.C. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res* **2016**, *5*, 2122, doi:10.12688/f1000research.9501.2.
35. Ahlmann-Eltze, C.; Huber, W. glmGamPoi: fitting Gamma-Poisson generalized linear models on single cell count data. *Bioinformatics* **2021**, *36*, 5701-5702, doi:10.1093/bioinformatics/btaa1009.
36. Van den Berge, K.; Perraudeau, F.; Soneson, C.; Love, M.I.; Risso, D.; Vert, J.P.; Robinson, M.D.; Dudoit, S.; Clement, L. Observation weights unlock bulk RNA-seq tools for zero inflation and single-cell applications. *Genome Biol* **2018**, *19*, 24, doi:10.1186/s13059-018-1406-4.
37. Zhu, A.; Ibrahim, J.G.; Love, M.I. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics* **2019**, *35*, 2084-2092, doi:10.1093/bioinformatics/bty895.
38. Hao, Y.; Hao, S.; Andersen-Nissen, E.; Mauck, W.M., 3rd; Zheng, S.; Butler, A.; Lee, M.J.; Wilk, A.J.; Darby, C.; Zager, M.; et al. Integrated analysis of multimodal single-cell data. *Cell* **2021**, *184*, 3573-3587 e3529, doi:10.1016/j.cell.2021.04.048.
39. Butler, A.; Hoffman, P.; Smibert, P.; Papalexi, E.; Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **2018**, *36*, 411-420, doi:10.1038/nbt.4096.
40. Stuart, T.; Butler, A.; Hoffman, P.; Hafemeister, C.; Papalexi, E.; Mauck, W.M., 3rd; Hao, Y.; Stoeckius, M.; Smibert, P.; Satija, R. Comprehensive Integration of Single-Cell Data. *Cell* **2019**, *177*, 1888-1902 e1821, doi:10.1016/j.cell.2019.05.031.